

# Low Delay Moving-Horizon Multiple-Description Audio Coding for Wireless Hearing Aids

Østergaard, Jan; E. Quevedo, Daniel; Jensen, Jesper

*Published in:*
I E E E International Conference on Acoustics, Speech and Signal Processing. Proceedings

*DOI (link to publication from Publisher):*
10.1109/ICASSP.2009.4959510

*Publication date:*
2009

*Document Version*
Early version, also known as pre-print

# LOW DELAY MOVING-HORIZON MULTIPLE-DESCRIPTION AUDIO CODING FOR WIRELESS HEARING AIDS

*Jan Østergaard* [1], *Daniel E. Quevedo* [2], *and Jesper Jensen* [3]

[1]Department of Electronic Systems, Aalborg University, Aalborg, Denmark
[2]School of Electrical Engineering and Computer Science, The University of Newcastle, Australia
[3]Oticon, Copenhagen, Denmark
janoe@ieee.org, dquevedo@ieee.org, jsj@oticon.dk

## ABSTRACT

In this work, we construct a novel scheme for efficient perceptual coding of audio for robust communication between encoders and wireless hearing aids. To limit the physical size of the hearing aids and to reduce power consumption and thereby increase the lifetime expectancy of the batteries, the hearing aids are constrained to be of low complexity. We therefore provide an asymmetric strategy where most of the computational load is placed at the encoding side. We make use of multiple-description coding. This combats possible erasures on the wireless link between the encoder and the hearing aids without introducing significant delay. Furthermore, we employ psychoacoustically optimized noise-shaping quantizers based on the moving-horizon principle, which exploits a finite prediction horizon.

***Index Terms***— Multiple-description coding, noise shaping, perceptual audio coding, low delay source coding

## 1. INTRODUCTION

The aim of this work is to encode and communicate audio from a remote encoder (e.g., cell phone, ipod, radio, tv, concert) over a wireless link to a pair of hearing aids.

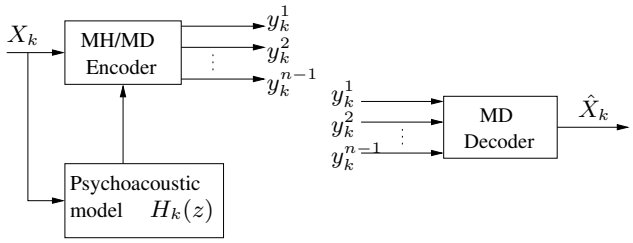
If the encoder is the hearing aid itself, a cell phone, or a tv, then it is essential that the latency is kept low. Low latency is important in order to establish lip synchronization, to avoid distortions due to a direct path acoustic signal reaching the eardrums out of synchronicity with the hearing aid output, and to facilitate a real-time communication situation We will assume that the maximum tolerable latency is on the order of a few milliseconds.

Due to battery and space considerations, the computational complexity at the decoder should be kept low. Thus, besides the cost of operating the antenna(s) and the demodulators, we only allow simple scaling and table look-up operations in this work.

Since the persons wearing the hearing aids are often not spatially stationary, the transmission channel is suscept to fading. In order to guarantee a certain degree of robustness towards channel impairments without introducing additional delay, we rely on multiple-description (MD) coding [1]. We consider the general case of $n$ channels. For example, each of the two hearing aids may have one or more receive antennas and furthermore, they communicate with each other. Thus, several channels are available even in the single person situation.

To achieve perceptually efficient encoding without introducing large delays, we employ moving-horizon (MH) quantization techniques at the encoder [2].

The work of Jan Østergaard is supported by the Danish Research Council for Technology and Production Sciences, grant no. 274-07-0383.




(a) Encoder (b) Decoder

**Fig. 1**. The encoder consists of two parts; the moving-horizon multiple-description *MH/MD Encoder* and the *Psychoacoustic model*.

MD coding was recently used for robust perceptual audio coding [3, 4, 5]. In [3, 4], the case of two descriptions was considered, whereas in [5] it was shown, that even with highly unreliable networks, it is possible to achieve audio streaming of acceptable quality by using more than two descriptions. In [4, 5], perceptual models were derived at the encoder. These needed to be encoded and transmitted to the decoder as side information, in addition to the encoded audio data. It turns out that the bit rate required for encoding the perceptual model is up to 8 kbps [4, 5]. Since this model is required in all the descriptions, the bit rate of the side information can be significant. Moreover, it is an open question how to optimally distribute the bit budget between the perceptual model and the actual audio data.

MH quantization was recently cast in the framework of low delay audio coding [2]. In [2], given a fixed perceptual weighting filter, it was shown that, by increasing the optimization horizon, better performance could be achieved at the expense of more complexity at the encoder. The delay of the design in [2], was dictated by that of the optimization horizon, i.e. was on the order of a few samples.

In the work presented in this paper, we first extend [2] to the case of a time-varying perceptual weighting filter. We then show how one can combine MD coding and MH quantization in a perceptually efficient manner. The overall delay of the proposed design, depends upon the choice of perceptual model. For example, if the psychoacoustic model of MPEG1 layer 1 [6] is chosen, then the delay is about 6 ms. at 44.1 kHz. sampling frequency. This delay can be reduced to less than 1 ms. if we do not time-align the perceptual model with the current input sample. In our design, we do not need to transmit the perceptual weighting filter as side information to the decoder. Thus, we avoid the issue of having to distribute the bits between the audio data and the perceptual model.

The encoder and decoder of our proposal are presented in Fig. 1(a) and Fig. 1(b), respectively.

## 2. PRELIMINARIES

In this section, we present background material on MH quantization, psychoacoustic modelling, and MD quantization. We furthermore show how to adapt and extend these concepts so that they are applicable within our framework.

### 2.1. Moving Horizon Quantization

In MH quantization, the current scalar sample $x_k \in \mathbb{R}$ is combined with $N-1$ future samples and quantized using a vector quantizer $\mathcal{Q}_k^N(\cdot)$ [2]. Thus, the input to the quantizer is the $N$-dimensional vector $\bar{x}_k = (x_k, x_{k+1}, \cdots, x_{k+N-1})$ and the output of the quantizer, i.e. the quantized version of $\bar{x}_k$ is the vector $\bar{y}_k = (y_k, y_{k+1}, \cdots, y_{k+N-1})$. More precisely, given the current input vector $\bar{x}_k$, the quantizer $\mathcal{Q}_k^N(\cdot)$ minimizes a cost function, $J_k^N(\cdot)$, which includes perceptual weighting. In this work, we define the cost function to be

$$J_k^N(\bar{x}_k) \triangleq \sum_{i=k}^{k+N-1} \epsilon_i^2 \tag{1}$$

where $\epsilon_i \in \mathbb{R}$ is the perceptually weighted error at the $i$th time-lag, that is

$$\epsilon_i \triangleq \bar{h}_i * (\bar{x} - \bar{y}) \triangleq \sum_{n=0}^{K} h_{i,n}(x_{i-n} - y_{i-n}) \tag{2}$$

where $\bar{h}_i = (h_{i,0}, h_{i,1}, \ldots, h_{i,K})$ denotes the set of filter coefficients of the perceptual weighting filter $H_i(z)$ to be used at time $i$ and $*$ is the linear convolution operator. Thus, $\epsilon_i(z) = H_i(z)(\bar{x}(z) - \bar{y}(z))$ and

$$H_i(z) = 1 + \sum_{n=1}^{K} h_{i,n} z^{-n} \tag{3}$$

is a causal linear time varying filter of finite order $K$. In (3), $h_{i,0} = 1$ for all $i$ and $h_{i,n} = 0$ for $n < 0$ and $n > K$.

It follows that, given an input vector $\bar{x}_k$, the (locally) optimal output vector $\bar{y}_k^* = \mathcal{Q}_k^N(\bar{x}_k)$ (for the current time $k$) is given by

$$\bar{y}_k^* = \arg\min_{\bar{y}_k \in \mathcal{Y}} J_k^N(\bar{x}_k) \tag{4}$$

where $\mathcal{Y}$ denotes the constrained alphabet (or codebook) of $\bar{y}_k$.

The output of the MH encoder is then simply taken to be $y_k$, i.e. the first sample of the quantized vector $\bar{y}_k^*$. Thus, an MH encoder consists of the non-linear map $\mathcal{Q}_k^N(\bar{x}_k) = \bar{y}_k^*$ which is followed by a function that simply picks out the scalar element $y_k$. At any time $k$, the MH encoder therefore takes as input the current sample $x_k$ (as well as $N-1$ future samples) and outputs a single sample $y_k$.

It was shown in [2] that for the special case of $N = 1$ and a fixed perceptual weighting filter, MH quantization is algebraically equivalent to noise-shaping quantization. Choosing $N > 1$ gives, in general, a lower weighted reconstruction MSE.

### 2.2. Psychoacoustic Model

The specific choice of psychoacoustic model is not essential for our design. We can, for example, choose the model from the MPEG1 layer 1 standard [6], which is based on a block of $M = 512$ samples.

The perceptual filter $\bar{h}_k$, to be used at time $k$, is based on a block of $M$ samples. This block can be time-aligned with the current sample by e.g., allowing a delay of $M/2$ samples. Alternatively, the end point of the block can be aligned with the current sample, so that the block contains the current sample $x_k$, as well as the past $M-1$ samples.

At every time instant we update the filter. Thus, the sequence of filters $\{\bar{h}_k\}$ for $k = 0, 1 \ldots$, may be seen as a single time varying filter. We note that, due to the high degree of overlap between consecutive blocks of $M$ samples, the filters $\bar{h}_k$ and $\bar{h}_{k+1}$ are likely to be very similar in a mean square sense.

In order to obtain the perceptual filter $\bar{h}_k$ of order $K$, we use an idea suggested by Schuller et al. [7]. Let $|\theta_k(f)|^2$ be the masked threshold for the $k$th block, and notice that we would like to find a filter with a transfer function that satifies $|H_k(f)|^2 \approx |\theta_k(f)|^{-2}$. If we use $|\theta_k(f)|^2$ as a short-term power spectrum, then the symmetric autocorrelation sequence $\{r_{k,i}\}, i = 0, \ldots, \frac{M}{2}$, is found simply as the inverse DFT of $|\theta_k(f)|^2$. The filter coefficients $h_{k,1}, \ldots, h_{k,K}$ are now easily found from $\{r_{k,i}\}$ by use of the Yule-Walker equations [8].

At startup, we do not have any samples available at the encoder. To obtain the block of $M$ samples, which is required to perform the frequency analysis leading to the psychoacoustic model, we choose to use a growing block size. This avoids introducing startup delays. At time zero ($k = 0$), we only have the current sample $x_0$ available. One sample is clearly not enough information to establish an accurate frequency domain representation for the psychoacoustic model. We will therefore employ a fixed (pre-computed) average psychoacoustic filter of order $K$ for the first few time instances, say up to time $j$. At time $j$, we have access to $j+1$ samples and we use these to obtain a, possibly crude, approximation of the psychoacoustic model. To avoid excessive variations of the filter, we use a *smoothed* version of the filter, e.g. we may use a weighted average between $\bar{h}_j$ and $\bar{h}_{j-1}$. At time $M$, we use the filters derived above.

An important difference to previous work see, e.g., [7, 5] is that, in our case, we do not need the perceptual filter at the decoder. We do therefore not need to worry about whether smoothing the filter coefficients yields unstable inverse filters. Furthermore, we do not need to encode and transmit the filter coefficients.

### 2.3. State-Space Interpretation

Since we are working with time varying filters it is convenient to formulate the problem in the state space domain.

An equivalent minimal state-space form for the filter $H_k(z)$ is, see, e.g., [9]

$$H_k(z) = 1 + C_k(zI - A)^{-1}B \tag{5}$$

where $A \in \mathbb{R}^{K\times K}$, $B \in \mathbb{R}^{K\times 1}$, and $C_k \in \mathbb{R}^{1\times K}$ are given by

$$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; C_k^T = \begin{bmatrix} h_{k,1} \\ h_{k,2} \\ \vdots \\ h_{k,K} \end{bmatrix} \tag{6}$$

and are related to the sequence of filters $\{\bar{h}_k\}$ through [9]

$$h_{k,n} = C_k A^{n-1} B, \quad n = 1, \ldots, K, \quad k = 0, \ldots . \tag{7}$$

With this, we can express the weighted error $\epsilon_k \in \mathbb{R}$ as given by (2) in state-space form, that is

$$\bar{z}_{k+1} = A\bar{z}_k + B(x_k - y_k) \tag{8}$$

$$\epsilon_k = C_k\bar{z}_k + (x_k - y_k) \tag{9}$$

where $\bar{z}_k \in \mathbb{R}^K$ is the system state vector given by

$$\bar{z}_k = [x_{k-1} - y_{k-1}, x_{k-2} - y_{k-2}, \ldots, x_{k-K} - y_{k-K}]^T. \tag{10}$$

Based on this state-space representation, it can be shown that the cost function (1) can be rewritten as [2]

$$J_k^M(\bar{x}_k) = \|\Psi_k(\bar{x}_k - \bar{y}_k) + \Gamma_k \bar{z}_k\|^2 \tag{11}$$

where $\Psi_k \in \mathbb{R}^{N\times N}$ is given by

$$\Psi_k = \begin{bmatrix} h_{k,0} & 0 & \cdots & 0 \\ h_{k,1} & h_{k,0} & 0 & \vdots \\ \vdots & & \ddots & 0 \\ h_{k,N-1} & \cdots & h_{k,1} & h_{k,0} \end{bmatrix} \tag{12}$$

and $\Gamma_k \in \mathbb{R}^{N\times K}$ satisfies

$$\Gamma_k = \left[C_k^T, (C_k A)^T, \ldots, (C_k A^{N-1})^T\right]^T. \tag{13}$$

### 2.4. Multiple-Description Coding

In this section, we review traditional MD coding. In MD coding a single source vector $\bar{x}_k$ is mapped to multiple output vectors $(\bar{y}_k^0, \bar{y}_k^1, \ldots, \bar{y}_k^{n-1})$, which are usually referred to as descriptions [1]. In the general case, we have $n \geq 1$ descriptions. The problem is to design the $n$ encoders $f_j : \bar{x}_k \mapsto \bar{y}_k^j \in \mathbb{R}^N, j = 0, \ldots, n-1$ and $2^n$ decoders $g_\ell : \{\bar{y}_k^j : j \in \ell\} \mapsto \hat{\bar{y}}_k^\ell \in \mathbb{R}^N, \ell \subseteq \{0, \ldots, n-1\}$.

For every time instance $k$, the $n$ current descriptions, i.e. $\{\bar{y}_k^0, \bar{y}_k^1, \ldots, \bar{y}_k^{n-1}\}$, are transmitted over $n$ channels so that description $j$, i.e. $\bar{y}_k^j$, is transmitted on the $j$th channel. At any time $k$, an arbitrary subset of the channels may break down. Which of the channels are currently working is not known to the encoder, but it is known to the decoder.

The optimization problem of the encoders and decoders can be cast into a Lagrangian framework, where the partial distortions due to reconstructing using subsets of the descriptions are individually weighted by a set of Lagrangian weights. Specifically, let $0 \leq \gamma_\ell \in \mathbb{R}$ be the non-negative weight for the subset of descriptions indexed by $\ell$ where $\ell \subseteq \{0, \ldots, n-1\}$. The aim is to minimize some weighted[1] cost, say $J$, where

$$J_k^N(\bar{x}_k) \triangleq \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell D_\ell \tag{14}$$

and where $D_\ell$ is the expected distortion due to reconstructing using the set of descriptions indexed by $\ell$. A simple distortion metric is the mean squared error (MSE) defined by

$$D_\ell \triangleq \mathbb{E}\|\bar{X}_k - \hat{\bar{Y}}_k^\ell\|^2 \tag{15}$$

where $\bar{X}_k$ and $\hat{\bar{Y}}_k^\ell$ are random vectors.

The traditional MD design problem is then to find a set of jointly optimal encoders $\{f_j : j = 0, \ldots, n-1\}$ and decoders $\{g_\ell : \ell \subseteq \{0, \ldots, n-1\}\}$ such that (14) is minimized. This minimization is subject to rate constraints on the individual description rates $R_j, j = 0, \ldots, n-1$.

In this work we will construct the MD coders by use of index-assignments and lattice vector quantization following the design presented in [10]. Index-assignment (IA) based MD quantization is a technique first proposed by Vaishampayan [11]. In IA-based MD quantization, the source vector $\bar{X}_k$ is first quantized by a central quantizer $\mathcal{Q}_c$ in order to obtain the central reconstruction vector $\lambda_c \in \mathbb{R}^N$, i.e. $\lambda_c = \mathcal{Q}_c(\bar{X}_k)$. At this point, a non-linear function $\alpha$ is applied on $\lambda_c$ in order to map $\lambda_c$ to $n$ descriptions, i.e. $\alpha(\lambda_c) = (\lambda_0, \ldots, \lambda_{n-1})$, where $\lambda_i \in \mathbb{R}^N$ denotes the codeword for the $i$th description.[2] If the quantizers in question are lattice vector quantizers then $\lambda_c$ as well as $\lambda_i, i = 0, \ldots, n-1$, are all points in different lattices, see [12] for details.

[1] The weights $\gamma_\ell$ may, for example, reflect successful decoding probabilities, i.e. the probability of receiving only the descriptions, which are indexed by $\ell$.

## 3. MH/MD CODER FOR WIRELESS APPLICATIONS

In this section we describe our proposal. It brings together the MD coding paradigm described in Section 2.4 with the MH quantizer described in Sections 2.1 and 2.3.

### 3.1. Encoder

We first note that an MD encoder outputs multiple descriptions whereas the MH quantizer $\mathcal{Q}_k^N(\cdot)$ previously defined gives only a single output. Furthermore, there is a feedback loop at the encoder, since past decisions affect the current decision through the system state vector $\bar{z}_k$, see (10). In order for this feedback loop to to be well defined at the encoder, we need to form a single output based on the $n$ descriptions. Towards that end, we define[3] (see also Footnote 1)

$$\tilde{\bar{y}}_k \triangleq \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell \hat{\bar{y}}_k^\ell \tag{16}$$

and take the output $\tilde{y}_k$ to be the first sample of $\tilde{\bar{y}}_k$. Thus, at the encoder, we feed back $\tilde{y}_k$ and the vector $\bar{z}_k$ previously given by (10) is now formed as

$$\bar{z}_k = [x_{k-1} - \tilde{y}_{k-1}, x_{k-2} - \tilde{y}_{k-2}, \ldots, x_{k-K} - \tilde{y}_{k-K}]^T. \tag{17}$$

In order to construct the MD quantizer, we will adopt an offline design where the cost function given by (14) and (15) is minimized. Furthermore, recall from Section 2.1, that in MH quantization only the first output sample of the MH quantizer is transmitted to the decoder. Thus, in our case, we need to transmit the first sample $y_k^j$ of the descriptions $\bar{y}_k^j, j = 0, \ldots, n-1$ to the decoder. When designing the MD quantizer, we therefore need to minimize (14) subject to entropy constraints on the discrete entropy of only the first sample of each of the descriptions.

For any given number of descriptions $n$, bit rates $\{R_j\}_{j=0}^{n-1}$, and weights $\{\gamma_\ell\}_{\ell\subseteq\{0,\ldots,n-1\}}$, we use the method presented in [10] in order to design the MD quantizers and index assignment function $\alpha$, see Section 2.4. To be able to decode, when receiving only the first sample of the $n$ descriptions, we construct the $N$-dimensional MD quantizer as a cascade of $N$ scalar MD quantizers. Then, in the online process, we take into account the perceptual weighting by minimizing (14) where $D_\ell$ is now given by

$$D_\ell \triangleq \|\Psi_k(\bar{x}_k - \hat{\bar{y}}_k^\ell) + \Gamma_k \bar{z}_k\|^2. \tag{18}$$

The optimal set of $n$ descriptions is the one that minimizes the cost (14) and (18). The first sample of each of these $n$ $N$-dimensional vectors is then entropy coded and transmitted to the decoder.

[2] The mapping $\alpha$, which is usually called an index-assignment function, is invertible. Thus, if all $n$ descriptions, $\lambda_0, \ldots, \lambda_{n-1}$ are received, then the central reconstruction $\lambda_c = \alpha^{-1}(\lambda_0, \ldots, \lambda_{n-1})$ can be obtained.

[3] We note that how to form the vector to be fed back at encoder is a non-trivial problem. This is partly due to the fact that the encoder does not know in advance which descriptions will be received at the decoder.

### 3.2. Decoder

At the decoder, we receive a set of $0 \leq m \leq n$ descriptions, which we first entropy decode and then reconstruct $\hat{y}_k^\ell$ using the decoding map $g'_\ell : \{y_k^j : j \in \ell\} \mapsto \hat{y}_k^\ell \in \mathbb{R}$, where $\ell \subseteq \{0, \ldots, n-1\}$ denotes the indices of the received descriptions.[4] In particular, the simple decoding rule where the reconstruction is given by the average of the received descriptions generally works well [10, 5]. Thus, when $0 < m < n$, we set $\hat{y}_k^\ell = \frac{1}{m}\sum_{j\in\ell} y_k^j$, whereas when $m = n$ we let $\hat{y}_k^\ell = \alpha^{-1}(y_k^0, \ldots, y_k^{n-1})$. When no descriptions are received, we set $\hat{y}_k^\ell = 0$.

Notice that the reconstruction $\hat{y}_k^\ell$ is designed to be a good representation of $x_k$ from a perceptual point of view and will thus, in general, not correspond to an MSE estimate.

## 4. SIMULATIONS

We consider the situation of $n = 1, 2$, and 3 descriptions and fix the total bit rate as $R_T = 4$ bits/sample. We let the side description rate $R_s$ be the same for all descriptions, i.e. $R_s = R_T/n$. For example, if $n = 2$, then we have $R_s = 2$ bits/sample, whereas for $n = 3$ we have $R_s = 1.33$ bits/sample. We assume the packet loss probabilities to be i.i.d. with probabilities $p_i = p, i = 0, \ldots, n-1$. Furthermore, we let $p = 1\%, 5\%$, and $10\%$ and let the weights $\{\gamma_\ell\}$ be given by the probability of receiving the given set of packets, e.g., for $n = 2$ we have $\gamma_0 = \gamma_1 = (1-p)p, \gamma_{0,1} = (1-p)(1-p)$ and $\gamma = p^2$.

We design an entropy-constrained scalar IA-based two-description MD quantizer following the approach given in [10]. When the horizon length $N$ is greater than one, we form a vector (product) MD quantizer by using the same scalar MD quantizer along each of the $N$ dimensions. We use the psychoacoustic model of MPEG1 layer 1 and use a model order of $K = 15$. The test signal is $5 \times 10^4$ samples from a piece of jazz music having a sampling frequency of 44.1 kHz.

The total weighted distortion, $D_T$, due to reconstructing using $\hat{y}_k^\ell \in \mathbb{R}$ at time $k$ (at the decoder), is given by

$$D_T = \sum_k \left| \sum_{i=0}^{K} h_{k,i}(x_{k-i} - \hat{y}_{k-i}^\ell) \right|^2 . \qquad (19)$$

We have included $D_T$ for the above example in Table 1. Rows 4–5, refer to forming the feedback variable $\tilde{\bar{y}}_k$ at the encoder as in (16), whereas in rows 6–7 we simply adopted $\tilde{\bar{y}}_k = \alpha^{-1}(\bar{y}_k^0, \ldots, \bar{y}_k^{n-1})$, i.e. the central reconstruction. In both cases, we minimize (18). For comparison, we also construct a single description (SD) scheme, and repeat the description $n-1$ times, see rows 1–3. This gives a total of $n$ identical descriptions, all at a bit rate of $R_T/n$ bits/sample. This is a simple but less efficient way to construct multiple descriptions.

We observe from Table 1 that the MD/MH optimized framework yields a lower weighted distortion than the SD/MH framework (also in the case where we repeat the descriptions in the SD-setup). We also observe that by increasing the horizon length, the distortion is not always reduced. It is possible that this apparent inconsistency is due to the fact that the encoder does not know the exact loss pattern experienced by the decoder. Thus, in the case of larger horizon lengths, the deterministic online optimization performed at the encoder, which is based on a feedback variable comprising a weighted sum of the multiple descriptions, is not adequately matched to the stochastic behavior of the packet losses observed by the decoder.

[4] Notice that $g'_\ell$ is generally different from standard MD decoding $g_\ell$, see Sec. 2.4. Indeed, $g'_\ell$ works on a set of $m = |\ell|$ scalar elements, whereas $g_\ell$ works on a set of $m$ $N$-dimensional vectors.

| | | $p = 1\%$ | | $p = 5\%$ | | $p = 10\%$ | |
|---|---|---|---|---|---|---|---|
| $n$ | Method | $N=1$ | $N=2$ | $N=1$ | $N=2$ | $N=1$ | $N=2$ |
| 1 | MH/SD | 76.56 | 77.57 | 326.29 | 326.04 | 635.87 | 633.60 |
| 2 | MH/SD | 89.80 | 88.55 | 114.80 | 119.52 | 209.64 | 206.83 |
| 3 | MH/SD | 219.48 | 216.66 | 220.09 | 218.49 | 232.83 | 230.72 |
| 2 | MH/MD | 49.71 | 41.75 | 103.62 | 107.90 | 201.85 | 202.43 |
| 3 | MH/MD | 31.99 | 30.53 | 108.01 | 103.56 | 200.29 | 196.33 |
| 2 | MH/MD | 44.25 | 44.20 | 106.36 | 107.98 | 201.65 | 198.48 |
| 3 | MH/MD | 31.67 | 31.43 | 106.20 | 105.16 | 202.67 | 197.69 |

**Table 1**. Total weighted distortions, see (19).

## 5. CONCLUSION

We presented a new idea in low delay perceptual audio coding for lossy networks. Specifically, we showed that it is possible to combine MH quantization with MD coding. The techniques complement each other well: The former makes it possible to take into account perceptual models when shaping the quantization noise, and the latter provides robustness towards packet losses.

## 6. REFERENCES


[1] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851 – 857, Nov. 1982.

[2] G. C. Goodwin and D. E. Quevedo, "Moving-horizon optimal quantizer for audio signals," *J. Audio Eng. Soc.*, vol. 51, no. 3, pp. 138 – 149, March 2003.

[3] R. Arean, J. Kovačević, and V. K. Goyal, "Multiple description perceptual audio coding with correlating transform," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 140 – 145, March 2000.

[4] G. Schuller, J. Kovačević, F. Masson, and V. K. Goyal, "Robust low-delay audio coding using multiple descriptions," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, Sep. 2005.

[5] J. Østergaard, O. A. Niamut, J. Jensen, and R. Heusdens, "Perceptual audio coding using $n$-channel lattice vector quantization," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2006, vol. 5, pp. 197 – 200.

[6] International Standard ISO/IEC 11172-3 (MPEG), "Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s. part 3: Audio," 1993.

[7] G. D. T. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *Trans. Speech and audio Proc.*, vol. 10, no. 6, pp. 379, Sep. 2002.

[8] J. D. Markel and A. H. Gray, *Linear prediction of speech*, Prentice Hall, 1976.

[9] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall, 1984.

[10] J. Østergaard, J. Jensen, and R. Heusdens, "$n$-channel entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1956 – 1973, 2006.

[11] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821 – 834, May 1993.

[12] J. Østergaard, *Multiple-description lattice vector quantization*, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, June 2007.